63-3-2

CU-36-63-Nonr-266(59) MS

401 803

COMBINATORIAL METHODS IN THE THEORY OF DAMS

by

Lajos Takács

Columbia University

Department of Mathematical Statistics

March 25, 1963

This research was supported by the Office of Naval Research under Contract Number Nonr-266(59), Project Number 042-205. Reproduction in whole or in part is permitted for any purpose of the United States Government.

1. INTRODUCTION

In this paper we shall be concerned with two mathematical models of infinite dams. In the first model independent random inputs occur at regular time intervals and in the second model independent random inputs occur in accordance with a Poisson process. The first model has already been studied by J. Gani [6], G. F. Yeo [16] and others, and the second model by J. Gani and N.U. Prabhu [7], J. Gani and R. Pyke [8], D. G. Kendall [9], and others. For both models we shall find explicit formulas for the distribution of the content of the dam and that of the lengths of the wet periods and dry periods. The proofs are elementary and based on two generalizations of the classical ballot theorem.

2. GENERALIZATIONS OF THE CLASSICAL BALLOT THEOREM

The classical ballot theorem is as follows: Suppose that in a ballot candidate A scores a votes and candidate B scores b votes. Let $a \geq \mu b$ where $\mu \geq 0$ is an integer. The probability that, throughout the counting, the number of votes for A is always greater than μ times the number of votes registered for B is

$$(1) \qquad P = \frac{a-\mu b}{a+b},$$

provided that all the voting records are equally probable.

Formula (1) for $\mu = 1$ was found in 1887 by J. Bertrand [4] and

for $\mu \geqq 1$, also in 1887, by É. Barbier [3]. The proof of (1) for $\mu = 1$ was given in 1887 by D. André [2] and for $\mu \geqq 1$ in 1924 by A. Aeppli [1].

The classical ballot theorem can also be formulated as follows: Let $v_r = 0$ if the r-th vote is cast for A and let $v_r = (\mu+1)$ if the r-th vote is cast for B. Let $n = a+b$ and $k = b(\mu+1)$. Then

$$(2) \quad P\left\{ v_1+\dots+ v_r < r \text{ for } r = 1,\dots,n \mid v_1+\dots+ v_n = k \right\} = 1 - \frac{k}{n},$$

if $0 \leqq k \leqq n$.

The author proved by mathematical induction that (2) also holds if, more generally, $v_1,\dots, v_n$ are interchangeable random variables assuming nonnegative integer values. (Cf. [10] and [11].) Moreover (2) also holds if $v_1,\dots, v_n$ are cyclically interchangeable random variables assuming nonnegative integer values. For this latter case a simple geometric proof was given by C. L. Mallows (oral communication). Cf. also J. C. Tanner [15] and M. Dwass [5]. In what follows we shall prove (2) for cyclically interchangeable random variables.

THEOREM 1. Let us suppose that $v_1, v_2,\dots, v_n$ are nonnegative, integer-valued random variables and that all the n cyclic permutations of $(v_1, v_2, \dots, v_n)$ have a common joint distribution. Then

(3) $P\{v_1+\dots+v_r < r \text{ for } r = 1,\dots, n \mid v_1+\dots+v_n = k\} = \begin{cases} 1 - \frac{k}{n} & \underline{\text{if}}\ 0 \leq k \leq n, \\ 0 & \underline{\text{otherwise,}} \end{cases}$

<u>provided that the left hand side is defined.</u>

PROOF. Let $k_1, k_2,\dots, k_n$ be fixed nonnegative integers with sum $k_1+ k_2+\dots+ k_n = k$ where $0<k<n$. Define $k_{j+n} = k_j$ for $j = 1,2,\dots$. We shall prove that among the n cyclic permutations of $(k_1, k_2,\dots, k_n)$ there are exactly $n-k$ for which the sum of the first r members is less than r for all $r = 1,2,\dots,n$. Hence the theorem immediately follows for $0<k<n$. If $k = 0$ or $k \geq n$, then (3) is trivially true.

Define $\Delta_j = j-(k_1 + \dots + k_j)$ for $j = 1,2,\dots$. Then $\Delta_{j+n} = \Delta_j + \Delta_n = \Delta_j + (n-k)$ for $j = 1,2,\dots$. Let m be the greatest positive integer for which $\Delta_m = \min(\Delta_1, \Delta_2,\dots, \Delta_n)$. Now we shall prove that there are exactly $n-k$ values among $i = m+1,\dots, m+n$ such that

(4) $$\Delta_i < \Delta_j \text{ for all } j = i+1,\dots, i+n,$$

that is, there are exactly $n-k$ permutations among $(k_{i+1},\dots,k_{i+n})$, $i = m+1,\dots, m+n$, for which the sum of the first r members is less than r for all $r = 1,2,\dots,n$.

Denote by $i_1, i_2,\dots,i_{n-k}$ the greatest indices such that $\Delta_{i_1} = \Delta_m+1$, $\Delta_{i_2} = \Delta_m+ 2,\dots, \Delta_{i_{n-k}} = \Delta_m + (n-k) = \Delta_{m+n}$ respectively. They exist because $\Delta_{j+1} - \Delta_j \leq 1$ for every j. By the definition of m we have $i_{n-k} = m+n$ and therefore $m < i_1 < i_2 <\dots< i_{n-k} = m+n$. Clearly (4) holds if and only if $i = i_1, i_2,\dots, i_{n-k}$. This proves the assertion.

Now we shall also prove a further generalization of the classical

ballot theorem. The following theorem was found by the author [12] for interchangeable random variables, however, we shall prove it here, slightly more generally, for cyclically interchangeable random variables.

THEOREM 2. Let us suppose that $\chi_1, \chi_2, \ldots, \chi_n$ are nonnegative random variables and that all the n cyclic permutations of $(\chi_1, \chi_2, \ldots, \chi_n)$ have a common joint distribution. Let $\tau_1, \tau_2, \ldots, \tau_n$ be the coordinates arranged in increasing order of n points distributed uniformly and independently of each other in the interval $(0, t)$. If $\{\chi_r\}$ and $\{\tau_r\}$ are independent sequences, then

$$(5) \quad P\{\chi_1 + \ldots + \chi_r \leq \tau_r \text{ for } r = 1, \ldots, n \mid \chi_1 + \ldots + \chi_n = y\} = \begin{cases} (1 - \frac{y}{t}) & \text{if } 0 \leq y \leq t, \\ 0 & \text{otherwise}, \end{cases}$$

provided that the left hand side is defined.

PROOF. Let

$$(6) \quad \chi(u) = \sum_{\tau_r \leq u} \chi_r$$

for $0 \leq u \leq t$. Then the left hand side of (5) can also be written as follows: $P\{\chi(u) \leq u \text{ for } 0 \leq u \leq t \mid \chi(t) = y\}$. Now define

$$(7) \quad v_r^{(m)} = \left[\frac{2^m}{t}\left(\chi\left(\frac{rt}{2^m}\right) - \chi\left(\frac{rt-t}{2^m}\right)\right)\right], \quad r = 1, 2, \ldots, 2^m,$$

where the symbol $[a]$ denotes the greatest integer $\leq a$. By Theorem 1 we have for $0 \leq y \leq t$ that

$$(8) \quad 1 - \frac{y}{t} \leq P\left\{v_1^{(m)} + \dots + v_r^{(m)} < r \text{ for } r = 1, \dots, 2^m \mid \chi(t) = y\right\} \leq 1 - \frac{y}{t} + \frac{n}{2^m}$$

because $v_i^{(m)}$, $i = 1, 2, \dots, 2^m$, are cyclically interchangeable random variables that assume nonnegative integer values and

$$\frac{2^m y}{t} - n \leq v_1^{(m)} + \dots + v_{2^m}^{(m)} \leq \frac{2^m y}{t}$$

if $\chi(t) = y$. If $m \to \infty$ in (8), then by the continuity theorem of probability we get that

$$(9) \quad P\{\chi(u) \leq u \text{ for } 0 \leq u \leq t \mid \chi(t) = y\} = 1 - \frac{y}{t}$$

for $0 \leq y \leq t$. This proves (5).

REMARK. In [13] we proved that if $v_1, v_2, \dots, v_n$ are interchangeable random variables that assume nonnegative integer values, then the probability that $v_1 + \dots + v_r < r$ holds for exactly j values among $r = 1, \dots, n$ given that $v_1 + \dots + v_n = k$ is

$$(10) \quad P_j = \sum_{i=n-j}^{k} \frac{(n-k-1)}{(i+1)(n-i-1)} P\{v_1 + \dots + v_{i+1} = i \mid v_1 + \dots + v_n = k\}$$

if $0 < k < n-1$ and $j = n-k, \dots, n-1$, and

$$(11) \quad P_j = \frac{1}{n}$$

if $k = n-1$ and $j = 1, 2, \dots, n$.

By using the same procedure as we used in proving Theorem 2 we can obtain from (10) and (11) the following result: If $\chi_1, \chi_2, \dots, \chi_n$

are nonnegative, interchangeable random variables, if $\chi(u)$ is defined by (6), and if $\rho(t)$ denotes the measure of the set $\{u: \chi(u) \leqq u$ and $0 \leqq u \leqq t\}$, then

$$(12) \quad P\{\rho(t) \leqq x \mid \chi(t) = y\} = \iint\limits_{\substack{t \leqq u+v \\ u \leqq y,\ v \leqq x}} \frac{1}{u}\left(\frac{t-y}{t-v}\right) P\{u < \chi(u) \leqq u+du \mid \chi(t) = y\}\, dv$$

if $y < t$ and $t-y \leqq x \leqq t$, and

$$(13) \quad P\{\rho(t) \leqq x \mid \chi(t) = t\} = \frac{x}{t}$$

if $0 \leqq x \leqq t$.

3. REGULAR INPUT

Suppose that at times $n = 1, 2, \ldots$ water of quantities $v_1, v_2, \ldots$ is flowing into a dam (reservoir) and the release is continuous at constant unit rate when the dam is not empty. Suppose that $v_1, v_2, \ldots, v_n, \ldots$ are identically distributed, mutually independent random variables that assume nonnegative integer values. Denote by η_n the content of the dam immediately after time n. Then we have

$$(14) \quad \eta_n = [\eta_{n-1} - 1]^+ + v_n\,, \quad n = 1, 2, \ldots,.$$

The initial content η_0 is a nonnegative integer. Denote by ρ_0 the

time of the first emptiness, i.e., the smallest value of n such that $\eta_n = 0$. Following the initial wet period (if any) dry periods and wet periods alternate. Denote by $\Theta_1, \Theta_2, \ldots, \Theta_n, \ldots$ the lengths of the successive wet periods other than the initial one. They are identically distributed, mutually independent random variables. The dry periods are also identically distributed, mutually independent random variables and independent of the wet periods. The probability that a dry period has length k is $[1-P\{v_1 = 0\}][P\{v_1 = 0\}]^{k-1}$ for $k = 1, 2, \ldots$.

By (14) we have

$$\eta_n = \max\{v_r + \ldots + v_n - (n-r) \text{ for } r = 1, \ldots, n \text{ and } \eta_0 + v_1 + \ldots + v_n - n\}. \tag{15}$$

In what follows we shall use the notation $N_0 = 0$ and $N_n = v_1 + \ldots + v_n$ for $n = 1, 2, \ldots$.

THEOREM 3. <u>We have</u>

$$P\{\eta_n \leq k \mid \eta_0 = i\} = P\{N_n \leq n+k-i\} - \sum_{j=1}^{n-i} \sum_{\ell=0}^{n-j-i} \left(1 - \frac{\ell}{n-j}\right) P\{N_j = j+k\} P\{N_{n-j} = \ell\}. \tag{16}$$

<u>and, in particular,</u>

$$P\{\eta_n = 0 \mid \eta_0 = i\} = \sum_{j=0}^{n-i} \left(1 - \frac{j}{n}\right) P\{N_n = j\}. \tag{17}$$

PROOF. If we replace $v_1, v_2, \ldots, v_n$ by $v_n, v_{n-1}, \ldots, v_1$ respectively in (15), then we obtain a new random variable

$$(18) \qquad \tilde{\eta}_n = \max \left\{ N_r - r + 1 \quad \text{for} \quad r = 1, \ldots, n \quad \text{and} \quad N_n - n + \eta_0 \right\} ,$$

which has the same distribution as η_r. Thus

$$(19) \qquad \begin{aligned} & P\left\{\eta_n \leqq k \mid \eta_0 = i\right\} = P\left\{N_n \leqq n+k-i\right\} - \\ & P\left\{N_n \leqq n+k-i \quad \text{and} \quad N_r \geqq r+k \quad \text{for some} \quad r = 1, \ldots, n\right\} . \end{aligned}$$

Let $r = j$ $(j = 1, \ldots, n-i)$ be the greatest r for which $N_r \geqq r+k$. Then $N_j = j+k$ and

$$(20) \qquad \begin{aligned} & P\left\{\eta_n \leqq k \mid \eta_0 = i\right\} = P\left\{N_n \leqq n+k-i\right\} - \sum_{j=1}^{n-i} P\left\{N_j = j+k\right\} . \\ & \qquad P\left\{N_n - N_j \leqq n - j - i \quad \text{and} \quad N_r - N_j < r - j \text{ for } r = j+1, \ldots, n\right\} . \end{aligned}$$

By Theorem 1

$$(21) \qquad \begin{aligned} & P\left\{N_r - N_j < r - j \quad \text{for} \quad r = j+1, \ldots, n \quad \text{and} \right. \\ & \left. N_n - N_j \leqq n+j-i\right\} = \sum_{\ell=0}^{n-j-i} \left(1 - \frac{\ell}{n-j}\right) P\left\{N_{n-j} = \ell\right\} . \end{aligned}$$

Putting (21) into (20) we get (16). If, in particular, $k = 0$,

then by Theorem 1

$$(22)\quad P\{\eta_n = 0 \mid \eta_0 = i\} = P\{N_n \leq n-i \text{ and } N_r < r \text{ for } r = 1, \ldots, n\} =$$

$$\sum_{j=0}^{n-i} P\{N_n = j\}\left(1 - \frac{j}{n}\right) .$$

THEOREM 4. <u>We have for</u> $i \geq 1$, <u>that</u>

$$(23)\quad P\{\theta_0 \leq n \mid \eta_0 = i\} = \sum_{j=i}^{n} \frac{i}{j} P\{N_j = j-i\} .$$

PROOF. Now we have

$$P\{\theta_0 \leq n \mid \eta_0 = i\} = P\{N_r = r-i \text{ for some } r = 1, \ldots, n\} =$$

$$(24)\quad \sum_{j=i}^{n} P\{N_j = j-i\} P\{N_j - N_r < j-r \text{ for } r = 1, \ldots, j-1 \mid N_j = j-i\}$$

and by Theorem 1 the second factor in the sum is equal to i/j.

THEOREM 5. <u>We have for</u> $i = 1, 2, \ldots$ <u>that</u>

$$(25)\quad P\{\theta_1 \leq n\} = \sum_{j=1}^{n} \frac{1}{j+1} P\{N_{j+1} = j \mid N_1 > 0\} .$$

PROOF. It can easily be seen that

$$P\{\theta_1 \leq n\} = P\{N_{r+1} < r+1 \text{ for some } r = 1, \ldots, n \mid N_1 > 0\} =$$

$$(26)\quad \sum_{j=1}^{n} P\{N_{j+1} = j \mid N_1 > 0\} P\{N_{j+1} - N_{r+1} < j-r \text{ for } r = 1, \ldots, j \mid N_{j+1} = j\}$$

and (25) follows from Theorem 1.

REMARK. If we suppose that the level of the dam may vary in the interval $(-\infty, \infty)$, that is, the dam never becomes empty, then the probability that in the time interval $(0, n)$ the total time during which the level is below the initial level given that $N_n = k$ is

$$(27) \quad P\{N_r < r \text{ for } j \text{ indices } r = 1, \ldots, n \mid N_n = k\} = \sum_{i=n-j}^{k} \frac{(n-k-1)}{(i+1)(n-i-1)} P\{N_{i+1} = i \mid N_n = k\}$$

if $0 < k < n-1$ and $j = n-k, \ldots, n-1$, and

$$(28) \qquad P\{N_r < r \text{ for } j \text{ indices } r = 1, \ldots, n \mid N_n = n-1\} = \frac{1}{n}$$

if $k = n-1$ and $j = 1, \ldots, n$. These follow from (10) and (11) respectively.

4. POISSON INPUT

Suppose that in the time interval $(0, \infty)$ water is flowing into a dam (reservoir) according to a random process and the release is continuous at constant unit rate when the dam is not empty. Denote by $\chi(t)$ the total quantity of water flowing into the dam during the time interval $(0, t)$. It is supposed that

$$(29) \qquad \chi(t) = \sum_{i=1}^{v(t)} \chi_i$$

where $\{v(t),\ 0 \leq t < \infty\}$ is a Poisson process of density λ and $\chi_1, \chi_2, \ldots, \chi_n, \ldots$ are identically distributed, mutually independent, positive random variables and independent of $\{v(t)\}$.

Denote by $\eta(t)$ the content of the dam at time t. The initial content $\eta(0) \geq 0$. Denote by θ_0 the time of the first emptiness; $\theta_0 = 0$ if $\eta(0) = 0$. Following the initial wet period (if any) dry periods and wet periods alternate. Denote by $\theta_1, \theta_2, \ldots, \theta_n, \ldots$ the lengths of the successive wet periods other than the initial one. They are identically distributed, mutually independent random variables. The dry periods are also identically distributed, mutually independent random variables and independent of the wet periods. The probability that a dry period has length $\leq x$ is $1 - e^{-\lambda x}$ for $x \geq 0$.

First we mention that by Theorem 2 we have

$$(30) \qquad P\{\chi(u) \leq u \ \text{ for } \ 0 \leq u \leq t \mid \chi(t) = y\} = (1 - \tfrac{y}{t})$$

if $0 \leq y \leq t$. For if we know that in the time interval $(0,t)$ there are $n > 0$ events in the Poisson process, then the occurrence times $\tau_1, \tau_2, \ldots, \tau_n$ have the same joint distribution as the coordinates arranged in increasing order of n points distributed uniformly and independently of each other in the interval $(0,t)$. Thus by (5) the probability that $\chi(u) \leq u$ for $0 \leq u \leq t$ given that $\chi(t) = y$ and $v(t) = n$ is $(1 - \frac{y}{t})$ if $0 \leq y \leq t$. Since this probability is independent

of the condition $\nu(t) = n$, (30) follows immediately.

By using (30) or the procedure which we used in proving Theorem 2 we obtain the following theorems corresponding to Theorems 3, 4, and 5. In what follows we shall use the notation $d_x P\{\chi(u) \leq x\} = P\{x < \chi(u) \leq x+dx\}$ regardless of whether u depends on x or not.

THEOREM 6. *If* $c \geq 0$ *and* $x \geq 0$, *then*

$$(31) \quad P\{\eta(t) \leq x \mid \eta(0) = c\} = P\{\chi(t) \leq t+x-c\} - \iint\limits_{\substack{u+v \leq t-c \\ 0 \leq u,\, 0 \leq v}} \left(1 - \frac{v}{t-u}\right) d_u P\{\chi(u) \leq u+x\}\, d_v P\{\chi(t-u) \leq v\} .$$

In particular,

$$(32) \quad P\{\eta(t) = 0 \mid \eta(0) = c\} = \int_0^{t-c} \left(1 - \frac{y}{t}\right) d_y P\{\chi(t) \leq y\}$$

if $t \geq c$ *and* 0 *otherwise.*

THEOREM 7. *If* $c > 0$, *then*

$$(33) \quad P\{\theta_0 \leq t \mid \eta(0) = c\} = \int_c^t \frac{c}{u}\, d_u P\{\chi(u) \leq u-c\}$$

if $t \geq c$, *and* 0 *if* $t < c$.

THEOREM 8. *If* $t \geqq 0$, *then*

$$(34) \qquad P\{\theta_i \leqq t\} = \frac{1}{\lambda}\int_0^t \frac{1}{u}\, d_u\, P\{0 < \chi(u) \leqq u\}$$

for $i = 1, 2, \ldots,$ *and* $\lambda = \lim_{t \to 0} P\{\chi(t) > 0\}/t.$

The above three theorems have been proved in [4] under the more general assumption that $\{\chi(t),\ 0 \leqq t < \infty\}$ is a stochastic process with nonnegative, stationary, independent increments.

REMARK. If we suppose that the level of the dam may vary in the interval $(-\infty, \infty)$, that is, the dam never becomes empty, then the probability that in the time interval $(0, t)$ the total time during which the level is below the initial level given that $\chi(t) = y$ is equal to (12) or (13) where now $\{\chi(t),\ 0 \leqq t < \infty\}$ is defined by (29).

REFERENCES

[1] A. Aeppli: Zur Theorie verketetter Wahrscheinlichkeiten. Thèse, Zürich, 1924.

[2] D. André: Solution directe du problème résolu par M. Bertrand. C. R. Acad. Sci. (Paris) 105 (1887) 436-437.

[3] É. Barbier: Généralisation du problème résolu par M. J. Bertrand. C. R. Acad. Sci. (Paris) 105 (1887) 407.

[4] J. Bertrand: Solution d'un problème. C. R. Acad. Sci. (Paris) 105 (1887) 369.

[5] M. Dwass: A fluctuation theorem for cyclic random variables. Ann. Math. Statistics 33 (1962) 1450-1454.

[6] J. Gani: Elementary methods for an occupancy problem of storage. Math. Annalen 136 (1958) 454-465.

[7] J. Gani and N.U. Prabhu: The time dependent solution for a storage model with Poisson input. Jour. Math. and Mech. 8 (1959) 653-663.

[8] J. Gani and R. Pyke: The content of a dam as the supremum of an infinitely divisible process. Jour. Math. and Mech. 9 (1960) 639-651.

[9] D. G. Kendall: Some problems in the theory of dams. Jour. Roy. Stat. Soc. Ser. B 19 (1957) 207-212.

[10] L. Takács: The probability law of the busy period for two types of queuing processes. Operations Res. 9 (1961) 402-407.

[11] L. Takács: A generalization of the ballot problem and its application in the theory of queues. Jour. Amer. Stat. Ass. 57 (1962) 327-337.

[12] L. Takács: The time dependence of a single-server queue with Poisson input and general service times. Ann. Math. Statistics 33 (1962) 1340-1348.

[13] L. Takács: The distribution of majority times in a ballot.

[14] L. Takács: The distribution of the content of a dam when the input process has stationary independent increments.

[15] J. C. Tanner: A derivation of the Borel distribution. Biometrika 48 (1961) 222-223.

[16] G. F. Yeo: The time dependent solution for an infinite dam with discrete additive inputs. Jour. Roy. Stat. Soc. Ser. B 23 (1961) 173-179.